Segment Routing for IPv6 Networks

Andrew Yang (andryang@cisco.com)

System Engineer

CCIE R&S #13173

Cisco live!

Agenda

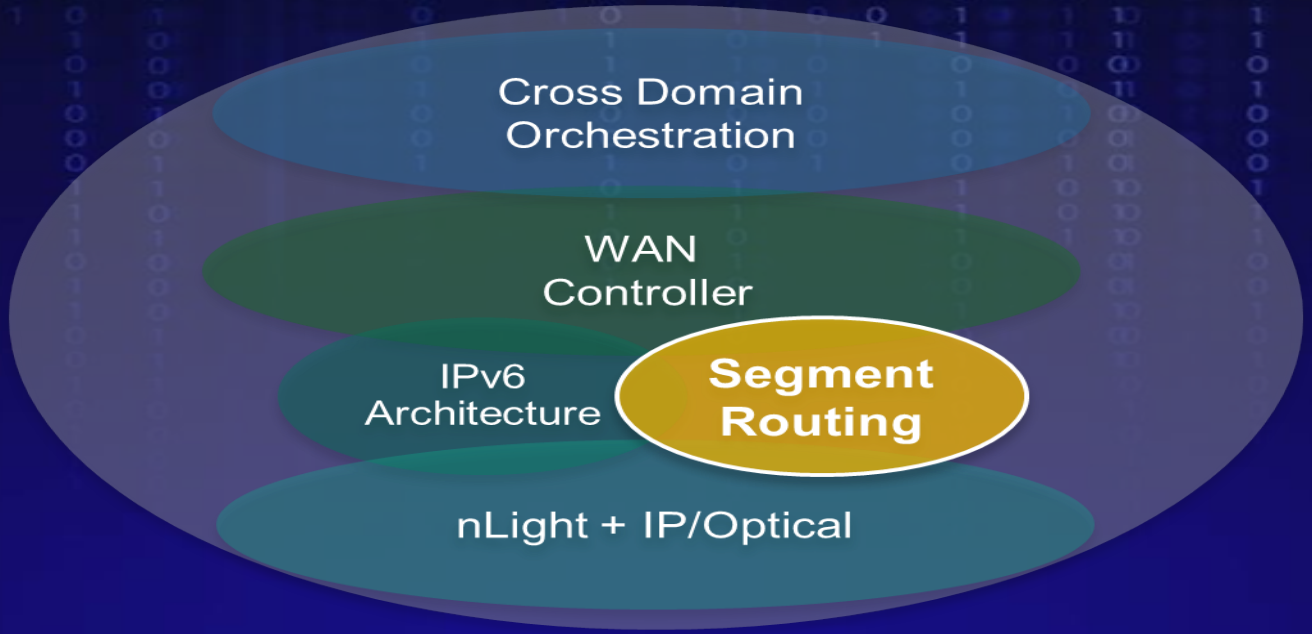
- Introduction to Segment Routing
- Segment Routing and the IPv6 Source Routing Model
- IPv6 Segment Routing Header (SRH)
- Use Cases

© 2015 Cisco and/or its affiliates. All rights reserved.

Introduction to Segment Routing

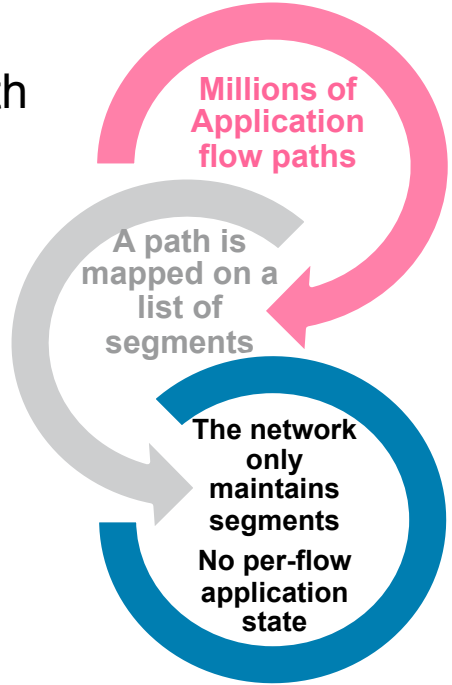
Cisco live!

How to achieve faster service velocity without increasing complexity and whilst managing costs?



Segment Routing – Scalability and Virtualization

- Each engineered application flow is mapped on a path
  - millions of paths
- A path is expressed as an ordered list of segments
- The network maintains segments
  - thousands of segments
  - completely independent of application size/frequency
- Excellent scaling and virtualization
  - the application state is no longer within the router but within the packet



© 2015 Cisco and/or its affiliates. All rights reserved.

Operators assume

- Leverage the classic IPv6 dataplane
  - only use the SR extension when needed
- Provide backward compatibility
  - incremental deployment
  - support non-SR intermediate nodes

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Operators ask for drastic improvement

- Provide a simple stateless IPv6 Traffic Engineering solution
  - less protocols to operate (avoid RSVP-TE)
  - avoid millions of tunnels to configure
  - avoid 100.000 of tunnel states on core routers
- Automated FRR for any topology

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Operators ask for a network model optimized for application interaction

- Applications must be able to interact with the network
  - cloud based delivery
  - internet of everything
- Programmatic interfaces and Orchestration
  - Necessary but not sufficient
- The network must respond to application interaction
  - Rapidly-changing application requirements
  - Virtualization
  - Guaranteed SLA and Network Efficiency

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing

- Simple to deploy and operate
  - Leverage IPv6 dataplane
  - Leverage Leancore hardware optimization
  - Straightforward ISIS/OSPF extension
- Stateless IPv6 Traffic Engineering solution
  - provide for optimum scalability, resiliency and virtualization
- Perfect integration with application
  - simple network, highly programmable
  - responsive to rapid changes

The state is no longer in the network but in the packet

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing

Simple

Segment = Instructions such as "go to node N using the shortest path"

- Source Routing: the source chooses a path and encodes it in the packet header as an ordered list of segments.
  - SR-IPv6: list of segment is encoded into a new (and secure) Routing Header
  - SR-MPLS: list of segments is represented by a label stack
- Segment: an identifier for any type of instruction
  - Service
  - Locator
  - IGP-based forwarding construct
  - BGP-based forwarding construct
  - Local value or Global Index
  - …

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public
Cisco live!

Segment Routing and the IPv6 Source Routing Model

Cisco live!

Segment Routing

- Forwarding state (segment) is established by IGP
  - Agnostic to forwarding dataplane: IPv6 or MPLS
  - RSVP-TE is not required
- Source Routing
  - source encodes path as an ordered list of segments
  - two segments: node or adjacency
- IPv6 SR Routing Extension header
  - includes the list of segments
- Tunneled solution
  - outer header protects the SR extension header – security
  - caches the active SID in the outer DA for backward compatibility

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

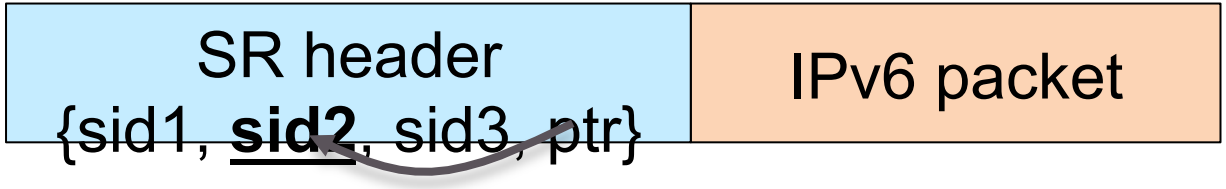
ECMP-aware Shortest Path

IPv6 packet

- The vast majority of the IP traffic travels on ECMP-aware shortest-path.
- SR header is obviously not needed for this case
- Plain classic IPv6 packet

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

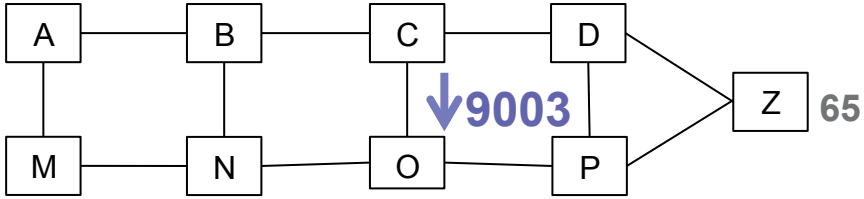
Source Routed Explicit Path



- SR header is present only for the subset of the traffic that requires source routed explicit path
- SR header contains
  - ordered list of segment
  - pointer identifying the active segment
  - forwarding decision is based on the active segment

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Adjacency Segment

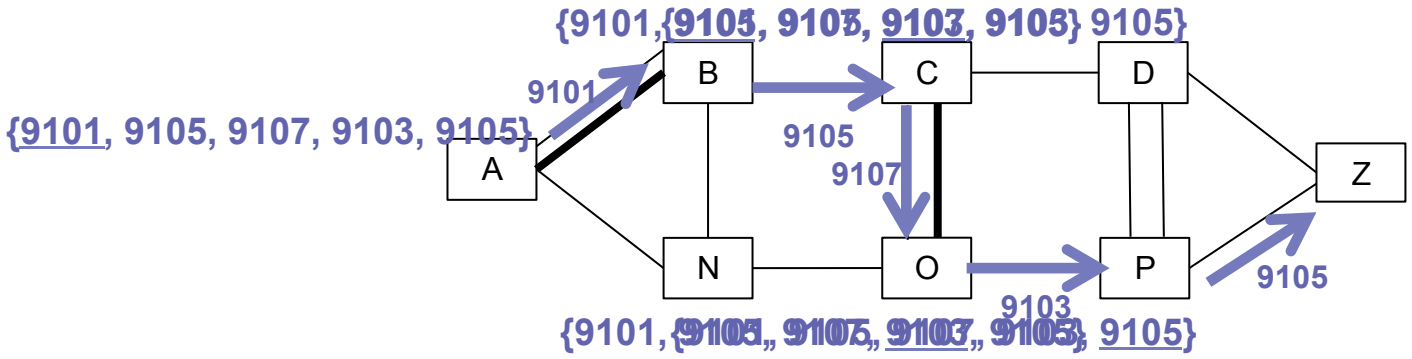


A packet injected at node C with active segment 9003 is forced through datalink CO

- C allocates a local segment for its datalink adjacency CO
- C advertises the adjacency segment in ISIS
  - simple sub-TLV extension
- C is the only node to install the adjacency segment in SR dataplane

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

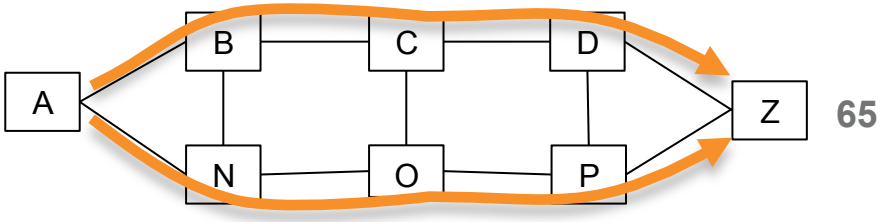
A path with Adjacency Segments



- Source routing along any explicit path
  - stack of adjacency segments
- SR provides for entire path control

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Node Segment

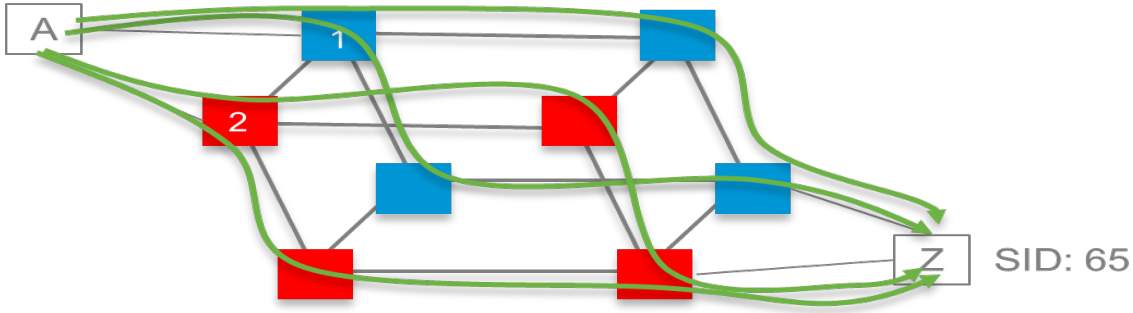


A packet injected anywhere with active segment 65 will reach Z via ecmp-aware shortest-path

- Z advertises its node segment
  - simple ISIS sub-TLV extension
- All remote nodes install the node segment to Z in the SR dataplane

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

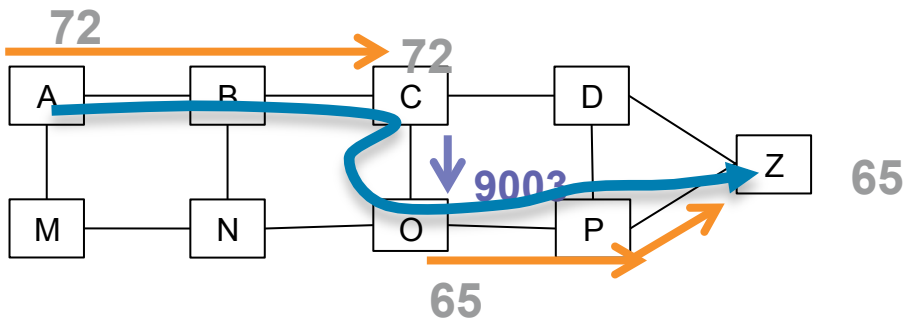
Node Segment



- ECMP-aware shortest-path to a node

© 2015 Cisco and/or its affiliates. All rights reserved.

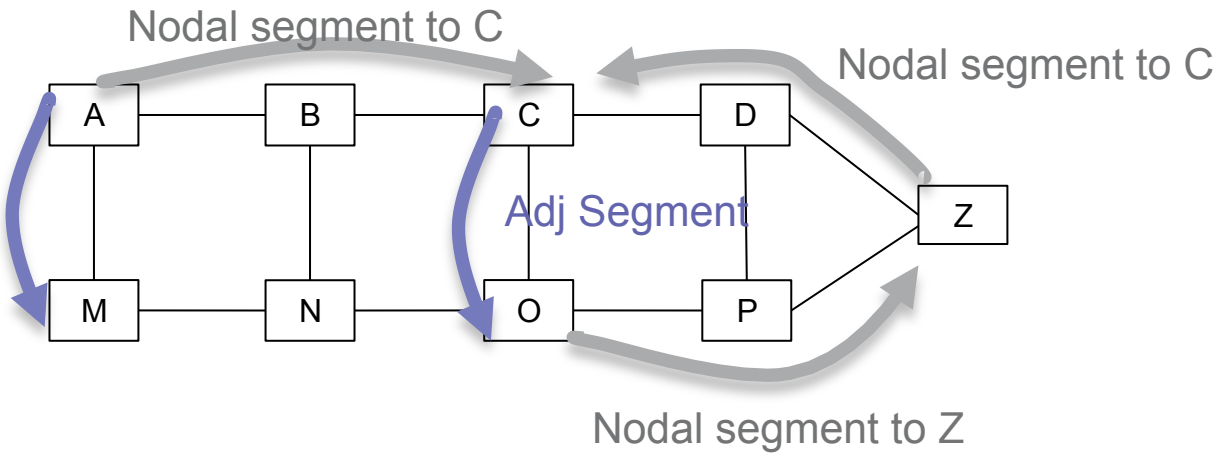
Combining Segments



- Source Routing
- ABCOPZ is expressed as {72, 9003, 65}

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

ISIS automatically installs segments

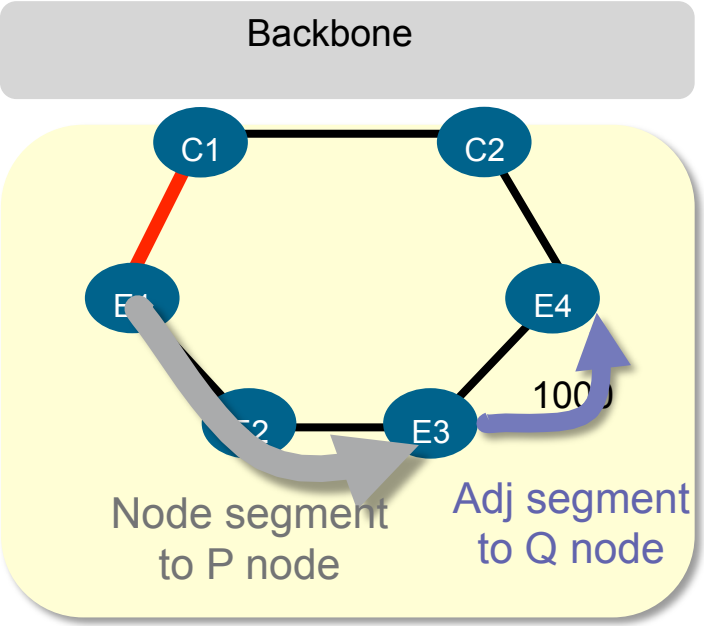


- Simple extension
- Excellent Scale: a node installs N+A FIB entries
  - N node segments and A adjacency segments

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Automated & Guaranteed FRR

- IP-based FRR is guaranted in any topology
  - 2002, LFA FRR project at Cisco
  - draft-bryant-ipfrr-tunnels-03.txt
- Directed LFA (DLFA) is guaranteed when metrics are symetric
- No extra computation (RLFA)
- Simple repair stack
  - node segment to P node
  - adjacency segment from P to Q



Default metric: 10

© 2015 Cisco and/or its affiliates. All rights reserved.

Segment Routing and the Source Based Routing Model Wait a Moment !!

- There are two ways of using Segment Routing on v6 networks
  - IPv6 control plane with a MPLS dataplane
  - IPv6 control plane with a IPv6 dataplane
- **This presentation covers Segment Routing for IPv6 control & data planes**

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing and the Source Based Routing Model

- SR-IPv6 allows IPv6 dataplane networks to benefit from all features deployed over the years on MPLS network:
  - Traffic Engineering
  - VPNs
  - Fast Reroute
  - Egress Peer Engineering
  - …

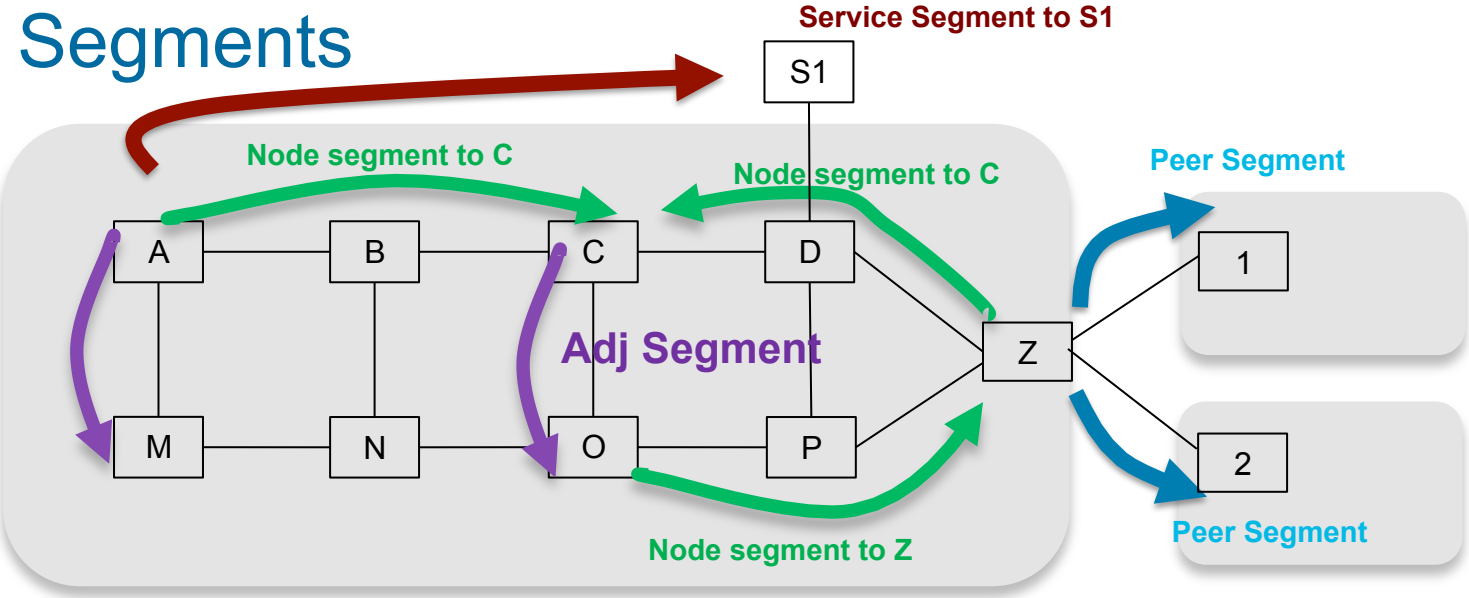
© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing IPv6 and the Source Based Routing Model

- Segment Routing:
  - Source based routing model where the source chooses a path and encodes it in the packet header as an ordered list of segments
  - A new type of the existing IPv6 Routing Extension Header is used for Segment Routing: SRH
  - The Segment Routing Header (SRH) contains the list of segments
    - Path state in the packet, not in the network
  - A segment is an instruction applied to the packet:
    - IGP-based forwarding construct
    - BGP-based forwarding construct
    - local adjacency
    - service/application
    - location,
    - context, …
  - Segment Routing leverages the source routing architecture defined in RFC2460 for IPv6

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Example of Segments



- Examples:
  - Go to this node using shortest path (Node-SID)
  - Go to this prefix using shortest path (Prefix-SID)
  - Go through this specific link (no matter what SPT says, Adj-SID)
  - Go through this egress interface / peering AS (Adj-SID, Peer-SID)
  - Etc.
- Simple protocol extensions allowing advertisement of segments
  - IGP, BGP, BGPLS, PCEP, …

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing and the Source Based Routing Model

- Segment Routing IPv6:
  - The notion of a “segment” is not new in IPv6
    - Routing Extension Header has been defined in RFC 2460 and defines the “segment”
    - In both RFC 2460 and Segment Routing a segment is identified by an IPv6 address
  - Segment Routing leverages RFC 2460 Routing Header by defining a new type
    - Improves Routing Header
    - Enhance the source routing model
    - Introduces security
  - Segment Routing does NOT require a forklift upgrade of the network
    - SR and non-SR nodes can co-exist
    - Gradual deployment
    - Full interoperability
    - Backward compatibility

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

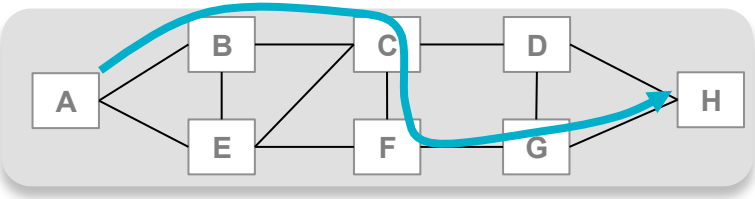
Segment Routing Model

- How to express an explicit (source routed) path knowing that:
  - Nodes may represent routers, hosts, servers, application instances, services, chains of services, etc.
  - A path is encoded into the packet by the originator (or ingress) node
  - A path may be modified by a node within the path
  - The network may have plurality of nodes not all supporting Segment Routing
  - A path can be “loose” or “strict”
    - Likely to be loose…

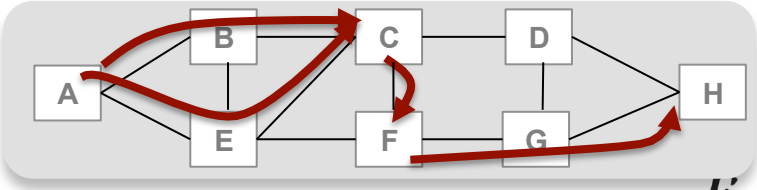
© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing Model

- Assuming following topology:
  - Node A has two shortest paths to C

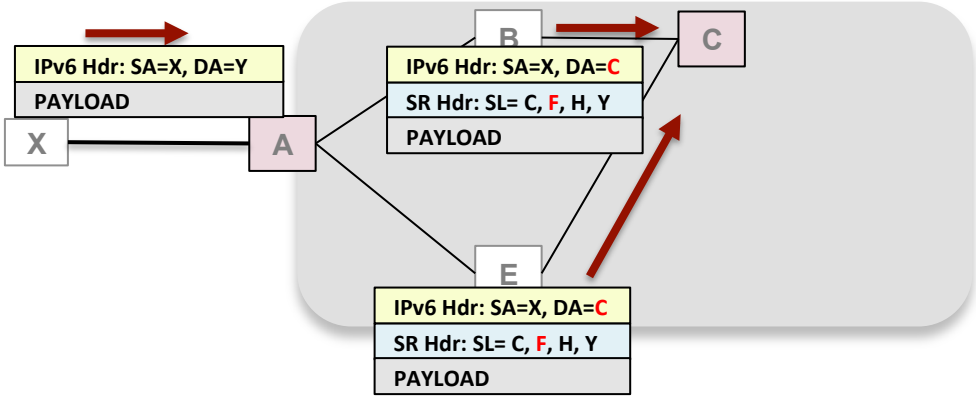


- Q: How to best express path: [A, B, C, F, G, H]
- A: Source rooted path with segments: [C,F,H]
  - >First segment: set of shortest paths from A to C (ECMP aware)
  - >Second segment: adjacency/link from C to F
  - >Third segment: shortest path from F to H
- Loose Source Routing



© 2015 Cisco and/or its affiliates. All rights reserved.

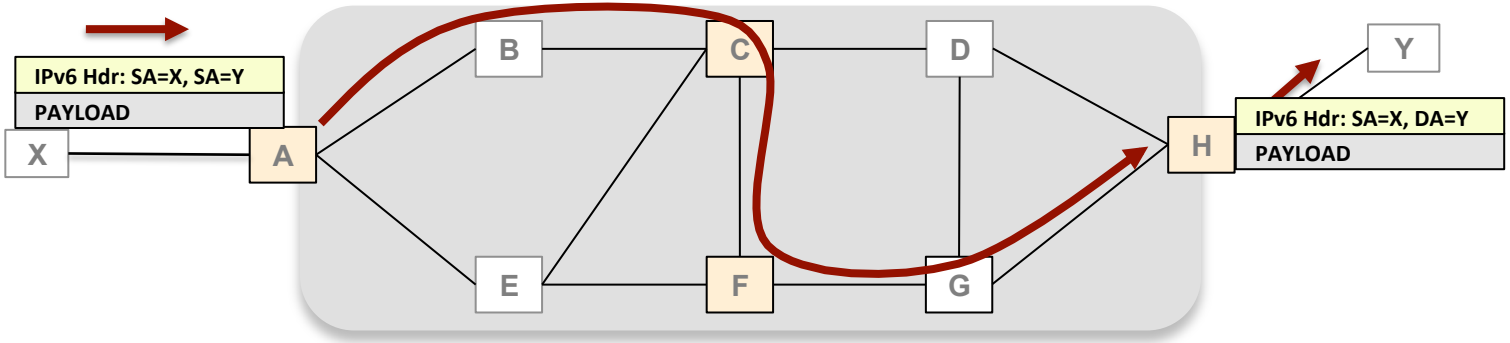
Segment Routing Header



- At ingress:
  - Path is computed or received by a controller (e.g.: SDN Controller)
  - Path is instantiated through a list of segments
  - A SRH is created with the segment list representing the path
  - Packet is sent to the first segment
    - >ECMP fully leveraged !

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

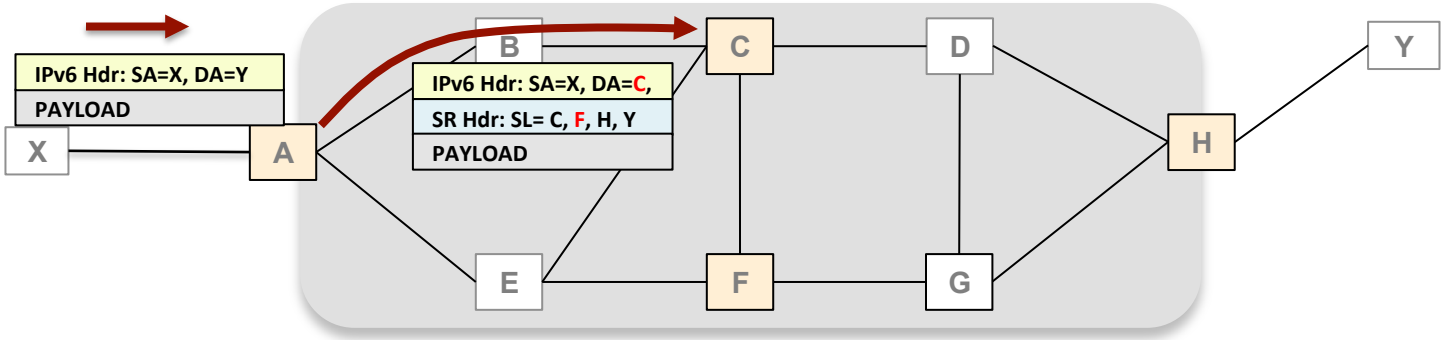
SR-IPv6 Example



- Example:
  - Classify packets coming from X and destined to Y and forward them across A,B,C,F,G,H path
  - Nodes A, C, F and H are SR capable
  - Nodes B, E, D and G are plain ipv6 forwarders

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

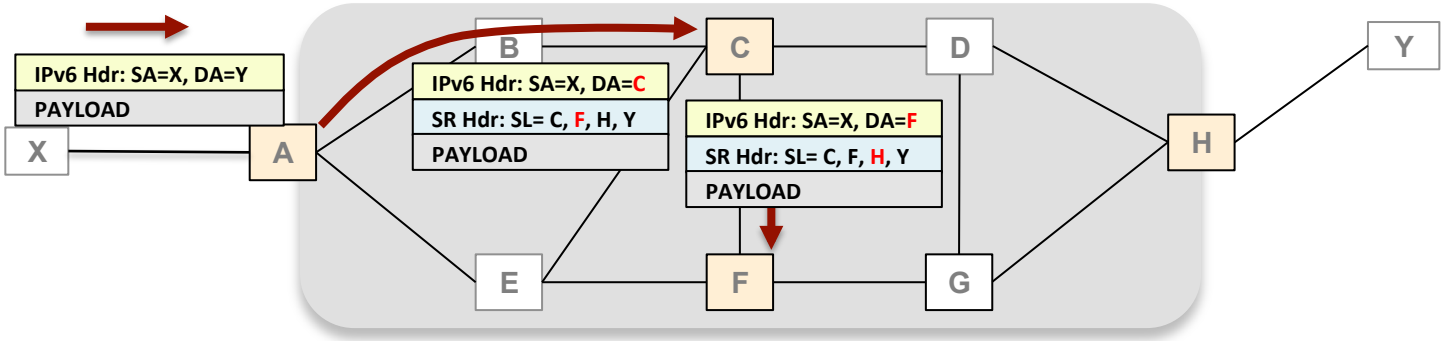
SR-IPv6 Example



- At ingress, the Segment Routing Header (SRH) contains
  - **Segment List**: C,F,H,Y (original destination address is encoded as last segment of the path)
  - **Segments Left:** points to the next segment of the path (F)
  - **DA** is set as the address of the first segment: C
- Packet is sent towards its DA (C, representing the first segment)
  - Packet can travel across non SR nodes who will just ignore the SRH
  - RFC2460 mandates only the node in the DA must examine the SRH

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

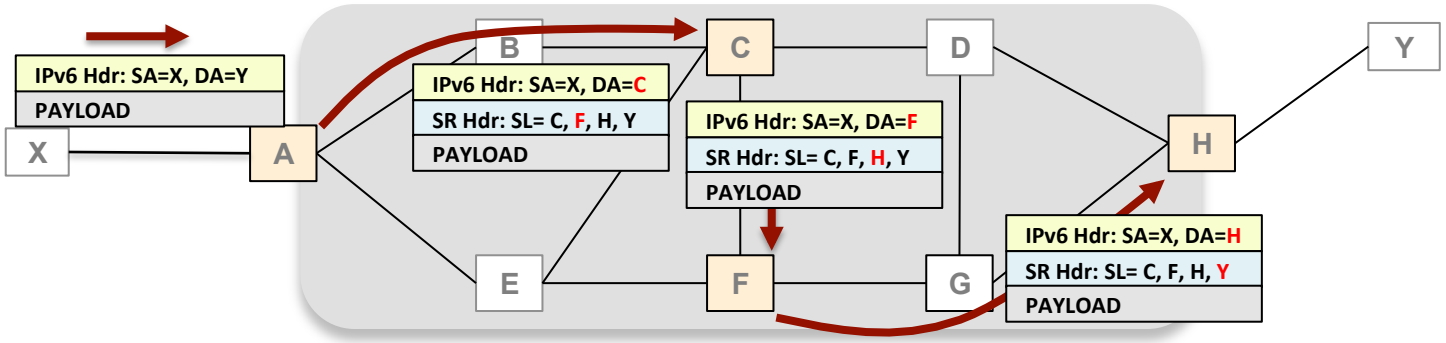
SR-IPv6 Example



- When packet reaches the segment endpoint C
  - Segment Left is inspected and used in order to update the DA with the next segment address: F
  - Segment Left pointer is decremented: now points to H
  - Packet is sent towards its DA

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

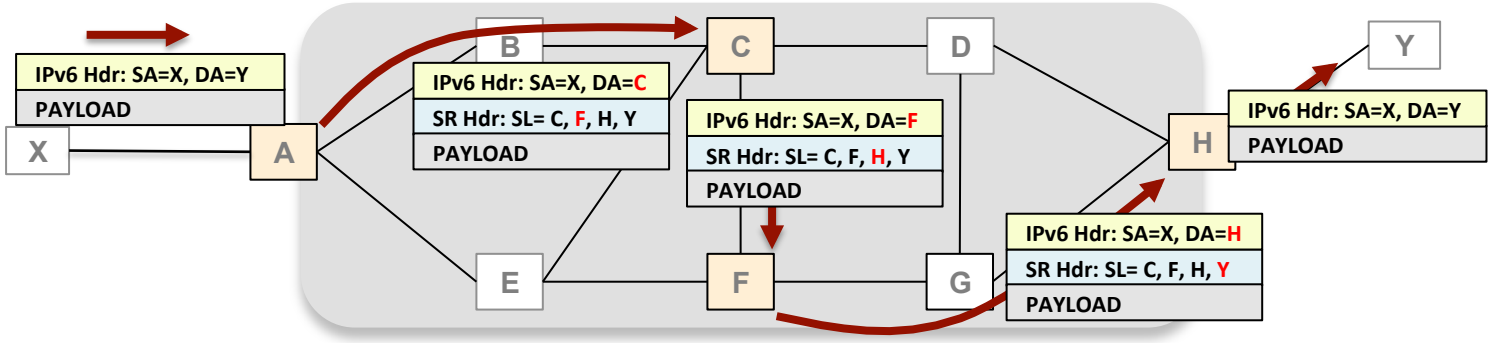
SR-IPv6 Example



- When packet reaches the segment endpoint F the same process is executed:
  - Segments Left is inspected and used in order to update the DA with the next segment address: H
  - Segments Left pointer is decremented: now points to Y (the original DA)
  - Packet is sent towards its DA

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

SR-IPv6 Example



- When packet reaches the segment endpoint H:
  - Segments Left is inspected and used in order to update the DA with the next segment address: Y
  - A flag (cleanup-flag) in SRH tells H to cleanup the packet and remove the SRH
  - Packet is sent towards its DA

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Segment Routing Header (SRH)

Cisco live!

SR Address

SRB-64bit-Block 0xCAFFE0123456789A	Active SID 32bits	Flags 8 bits	Entropy 24bits

- SR domain dedicates an /64 block for SR
  - called the SR block SRB (e.g. 0xCAFFE0123456789A/64)
  - other block length would work
- All nodes within domain are configured with SRB block
- The SID is found within the address
- Backward compatibility
  - Intermediate Non-SR-capable nodes switch the traffic classically

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

SR Forwarding

SRB-64bit-Block 0xCAFFE0123456789A	Active SID 32bits	Flags 8 bits	Entropy 24bits

- Upon receiving a packet, an SR node checks whether the DA is within the SRB block
- If yes, it forwards on the 32-bit SID
  - leancore performance improvement
  - it uses the 24-bit entropy to feed ECMP selection
    - ECMP efficiency is preserved despite the outer tunnel

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Stateless Traffic Engineering

- The operator or the SDN controller programs the ingress node to classify a specific flow and encapulate it with the SR source route
- The core of the network is not aware of anything
- A core node only installs
  - N node segments
  - A local adjacency segments
- Any route can be expressed based on these N+A segments
- Millions of flows can be engineered
- Frequent re-optimizations is possible

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

SRH (RH Type)

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Next Header   |  Hdr Ext Len  | Routing Type  | Segments Left |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| First Segment |               Flags           |  HMAC Key ID  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Segment List[0] (128 bits ipv6 address)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
                              . . .
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Segment List[n] (128 bits ipv6 address)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Policy List[0] (128 bits ipv6 address)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Policy List[1] (128 bits ipv6 address)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Policy List[2] (128 bits ipv6 address)             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            HMAC (256 bits, optional)                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Segment Routing Header:
  - **Segment List** describes the path of the packet: list of segments (IPv6 addresses)
  - **Segments Left**: Defined in [RFC2460], it contains the index, in the Segment List, of the next segment to inspect. Segments Left is decremented at each segment and it is used as an index in the segment list
  - **HMAC**
  - **Flags** and optional policy information
- The Active Segment is set as the DA of the packet
  - At each segment endpoint, the DA is updated with the next active segment found in the segment list
  - Compliant with RFC2460 rules for the Routing Header

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

SRH (RH Type TBD)

- Next Header: 8-bit selector. Identifies the type of header immediately following the SRH
- Hdr Ext Len: 8-bit unsigned integer. Defines the length of the SRH header in 8-octet units, not including the first 8 octets
- Type: TBD (SRH)
- Segments Left: index, in the Segment List, of the current active segment in the SRH. Decremented at each segment endpoint.
- First Segment: offset in the SRH, not including the first 8 octets and expressed in 16-octet units, pointing to the last element of the Segment List (i.e.: that contains the first segment of the path).
- Flags: 16 bits of flags. Following flags are defined:
  - bit-0: cleanup
  - bit-1: rerouted packet
  - bits 2 and 3: reserved
  - bits 4 to 15: policy flags. Define the type of the IPv6 addresses encoded into the Policy List (each address is described by 3 bits):
    - 0x0: Not present
    - 0x1: ingress SR PE address
    - 0x2: egress SR PE address
    - 0x3: original source address

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Next Header   |  Hdr Ext Len  | Routing Type  | Segments Left |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| First Segment |               Flags           |  HMAC Key ID  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Segment List[0] (128 bits ipv6 address)            |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
                           ...                                 .
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Segment List[n] (128 bits ipv6 address)            |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Policy List[0] (128 bits ipv6 address)             |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Policy List[1] (128 bits ipv6 address)             |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Policy List[2] (128 bits ipv6 address)             |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                            HMAC                            //
//                   (256 bits, optional)                     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```


© 2015 Cisco and/or its affiliates. All rights reserved.

SRH (RH Type TBD)

- Segment List[n]: 128 bit IPv6 addresses representing each segment of the path. The segment list is encoded in the reverse order of the path: the last segment is in the first position of the list and the first segment is in the last position.
- Policy List[n]: Addresses representing specific nodes in the SR path:
  Ingress SR PE: IPv6 address representing the SR node which has imposed the SRH (SR domain ingress)

  Egress SR PE: IPv6 address representing the egress SR domain node

  Original Source Address: IPv6 address originally present in the SA field of the packet
- HMAC: SRH authentication (optional)

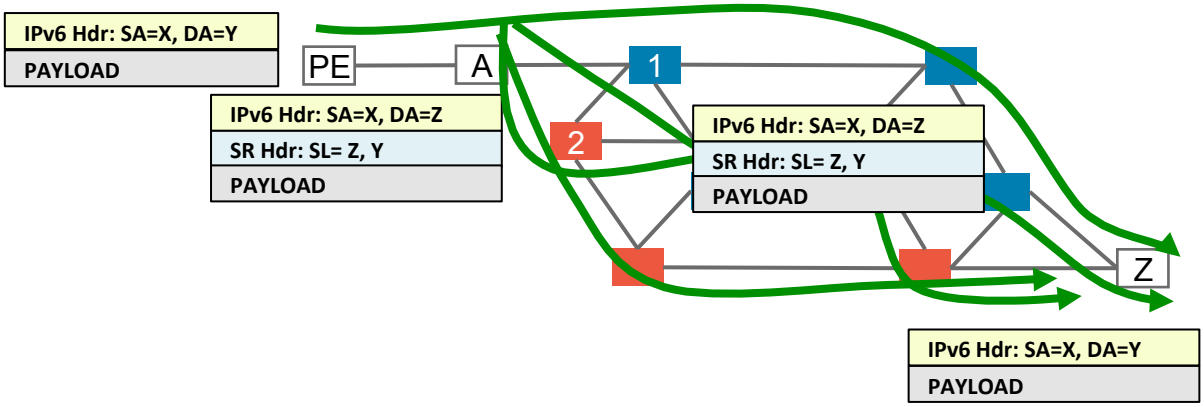
```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Next Header   |  Hdr Ext Len  | Routing Type  | Segments Left |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| First Segment |               Flags           |  HMAC Key ID  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Segment List[0] (128 bits ipv6 address)            |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
                            ...                                 .
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Segment List[n] (128 bits ipv6 address)            |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Policy List[0] (128 bits ipv6 address)             |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Policy List[1] (128 bits ipv6 address)             |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|            Policy List[2] (128 bits ipv6 address)             |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                            HMAC                            //
//                   (256 bits, optional)                     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```


© 2015 Cisco and/or its affiliates. All rights reserved.

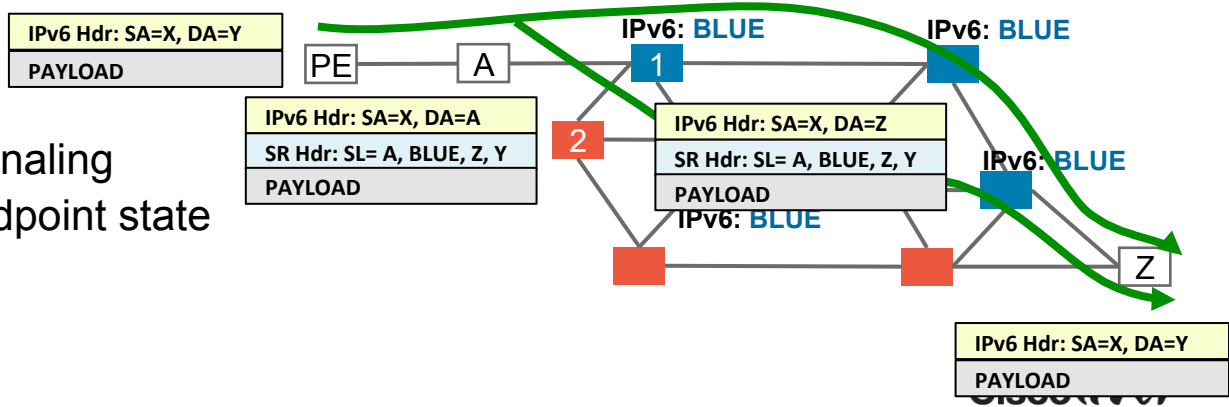
Segment Routing Use Cases

Cisco live!

Disjoint TE Service



- A to Z any plane
  - Set of IGP shortest paths to Z
- A to Z via blue plane
  - Traffic Engineering policy:
    - Set of shortest paths to BLUE (anycast IPv6 address/segment)
    - Set of shortest paths to Z
- Benefits
  - ECMP
  - Traffic Engineering with no signaling
  - Traffic Engineering with no midpoint state



© 2015 Cisco and/or its affiliates. All rights reserved.

Latency TE Service

- Data from Tokyo to Brussels
  - IGP shortest-path via US, higher and cheaper capacity
  - Prefix-SID of Brussels
- Voice from Tokyo to Brussels
  - SR TE policy uses one additional segment “Russia Anycast”
  - Low-latency path
- Benefits
  - ECMP
  - Availability of the anycast segment against node failure
  - No hop-by-hop signaling load and delay
  - No midpoint state

Node segment to Brussels

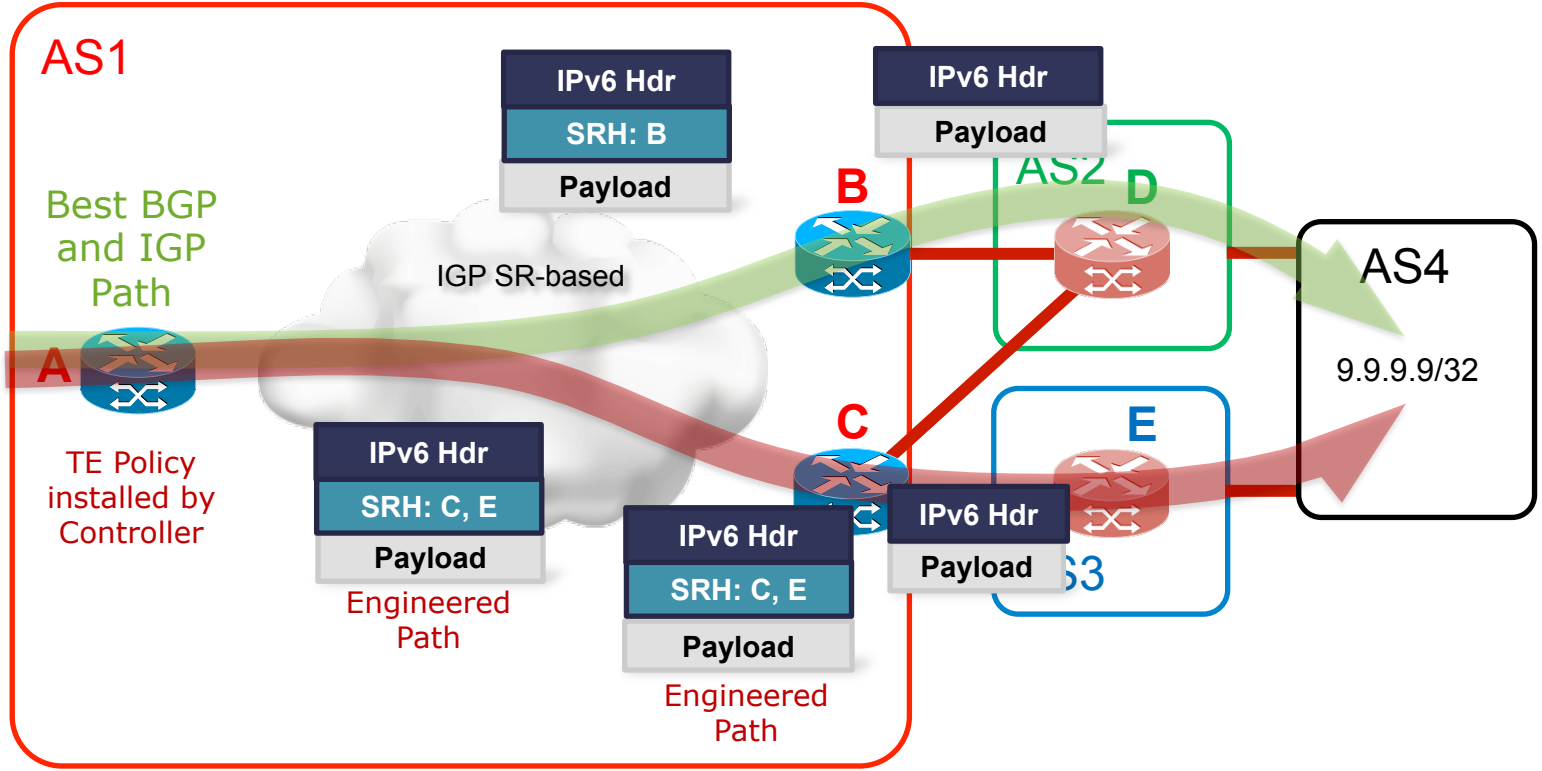
Node segment to Russia

IPv6 Hdr
SRH: Brussels
Data pkt

IPv6 Hdr
SRH: Russia Brussels
Voice pkt

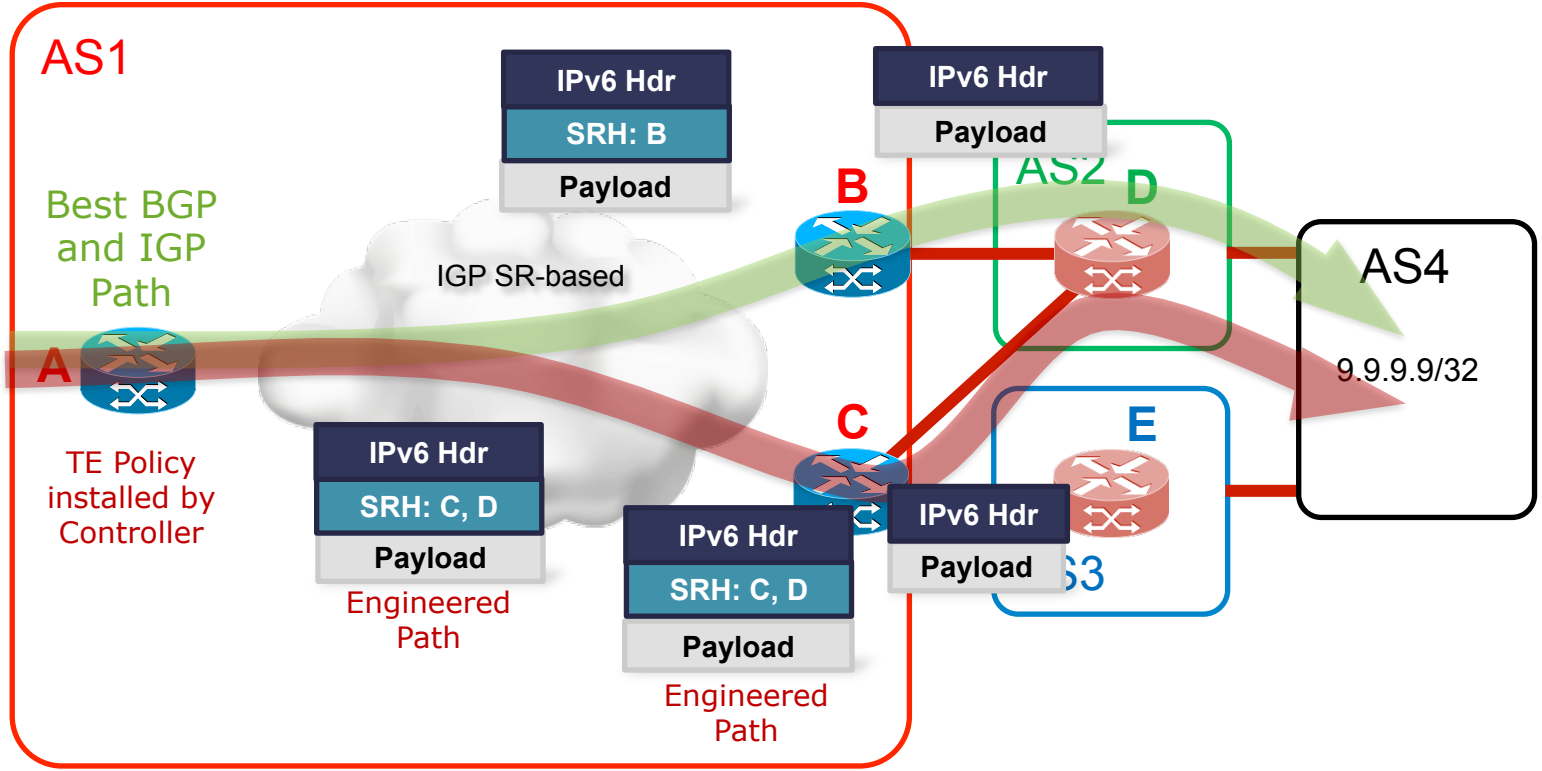
© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Public

Content producer engineers its WAN traffic to egress peers



© 2015 Cisco and/or its affiliates. All rights reserved.
Cisco Public

Content producer engineers its WAN traffic to egress peers



© 2015 Cisco and/or its affiliates. All rights reserved.
Cisco Public

Thank you.
cisco
Cisco live!